New Products & Services

In the Contour Reality System, the actor is captured by two arrays of synchronized digital cameras and lit by customized fluorescent lamps that strobe 90 to 120 times per second, beyond the threshold of human perception.

Contour Reality Capture

by David E. Williams

Computer-animated feature films involve the cinematography disciplines of composition, lighting and angles, and now even depth of field is being used to great effect. However, with rare exceptions, virtual cinematography is a specialty performed by those already working in that sphere, and not one where traditional directors of photography have had much opportunity to bring their talents into play. Unfortunately, asking a cinematographer to stay on as a paid post collaborator on visual-effects-heavy films is an infrequent request at best.

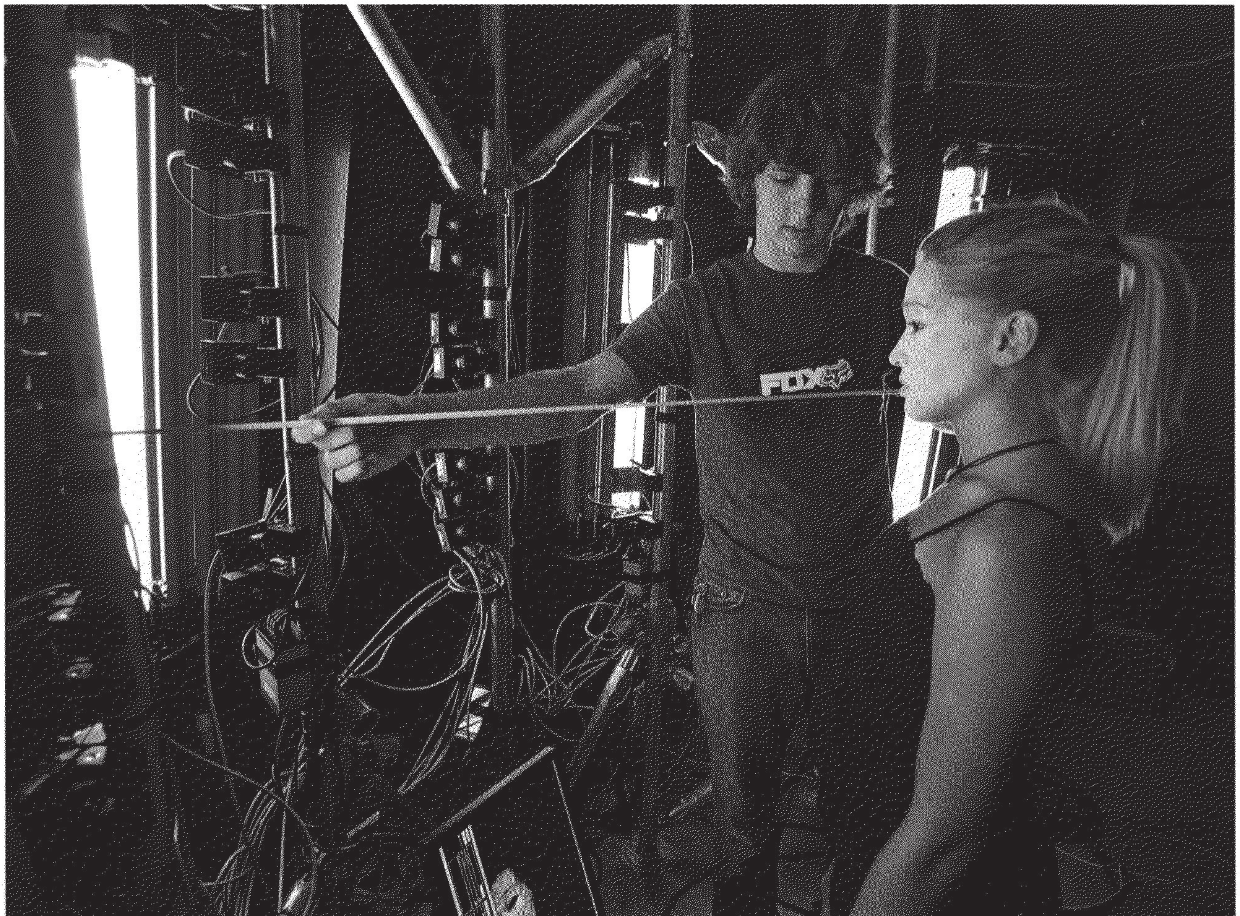
A keen eye for visual aesthetics and storytelling can make the most of the real and virtual filmmaking worlds, but the divide between them is crossed by only the flimsiest of technological and methodological bridges. One is marker-based motion capture, which records the movement of a human subject as a series of dots. Originally developed for applications such as sports medicine, mo-cap precisely tracks human skeletal motion by forming a 3-D stick figure of lines connected between the dots. But any onscreen character must be far more than a moving stick figure, so although mo-cap requires actors and a sense of space and timing, the collection of mo-cap data alone is not considered cinematography; instead, it is used as a starting point. But the keyframe-animation process required to flesh that stick figure out into a realistic character is expensive and time consuming. And in the case of creating CG humans, it is a technique that many believe has not realized its full potential. (Film critics often deride the resultant avatars as "lifeless" and "zombie-like.")

However, the Contour Reality Capture System, which was introduced at this year's Siggraph convention, may not only help bridge the traditional and

Preparing a performer for a Contour capture session requires just a few minutes of phosphorescent makeup application. The FDA-approved phosphorescent makeup can be mixed with base makeup to provide natural skin color.

virtual filmmaking worlds, but also breathe genuine life into "synthespian" performers. The system, a proprietary apparatus developed and now offered by the San Francisco mo-cap studio Mova, employs two synchronized camera systems to simultaneously capture 3-D geometric and full-color visual information of the subject. These two sets of data are later combined to result in a high-definition, volumetric, digital representation of the action that can be later imported, modified, manipulated or retargeted to a CG character using off-the-shelf animation software.

"Production tools exist today that give a director complete photorealistic control of every object within the 3-D volume of a scene, including camera position, composition, lighting, characters and props," says Steve Perlman, founder and CEO of the Palo Alto-based Rearden Companies, which owns and



Preparing to film with Contour. The custom fluorescent lamps serve to both recharge the phosphorescent makeup and provide dramatic subject lighting.

operates Mova. "But while we have powerful 3-D 'editing' tools, what we lack is a practical 3-D 'camera' that can shoot a live scene volumetrically with production-level quality. That's where Contour Reality Capture fits in. It's a volumetric cinematography system that captures all of the visible surfaces of a scene in 3-D."

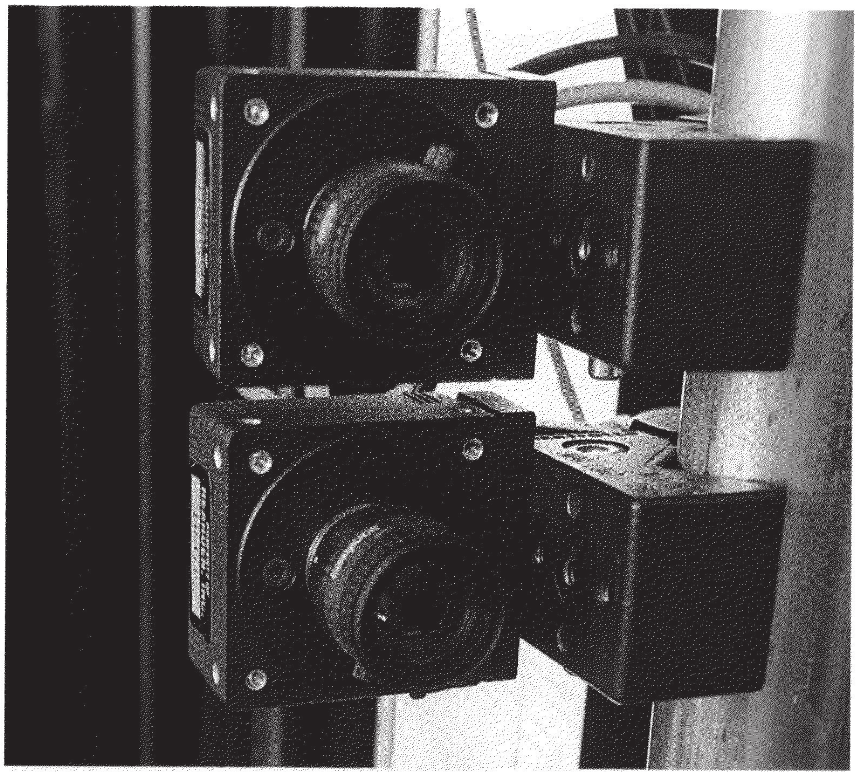
Mova was founded in 2004 by Rearden to provide 3-D mo-cap services using its Vicon MX-40 marker-based system, and the company's credits include the video games *The Godfather, From Russia With Love* and *Eragon.* Mova's sister company, Ice Blink Studios, which Perlman co-founded with Doug Chiang (production designer on *The Polar Express*), also is closely tied to mo-cap production, having provided visual effects and art direction for Sony Pictures' *Monster House* and Warner Bros.' upcoming *Beowulf.*

The chief architect behind Contour, Perlman is the holder of more than 60 patents pertaining to multimedia and communications technologies. He first attracted notice as a key researcher at Apple Computer, where he led multimedia initiatives, including the development of QuickTime, in the late 1980s. He is perhaps best known for founding WebTV Networks, which was acquired by Microsoft in 1997 and has since evolved into Microsoft's IPTV technology.

What exactly is "volumetric cinematography?" Perlman explains, "We all got a sneak preview of it in *The Matrix*, when Trinity [played by Carrie-Ann Moss] froze in mid-jump and the camera view spun around her. Despite having a large ring of cameras surrounding Moss, the camera motion was limited to a single path while the entire scene was frozen. Contour allows for the same level of realism achieved in that shot, but with the entire scene in motion, with complete flexibility of camera position, and with full control over lighting and compositing. And if the actress doesn't quite achieve the desired pose in her jump, you can use Contour to reposition her limbs in 3-D during post."

Traditionally, human skeletal dynamics have been recorded and represented in the form of mo-cap data, which is later combined with digital skin, hair and clothing. Although this process is well suited for capturing skeletal motion for CG characters, it captures only the barest clues of the motion of deformable surfaces, such as a human face, where we typically see the finer points of an actor's performance. Using conventional mo-cap to record the subtleties of a smile or furrowed brow is akin to the actor performing while wrapped in a head-to-toe latex suit, with their expressive eyes, fine facial characteristics and nuanced surface textures largely erased.

Conversely, Contour instantaneously records an entire human performance — simultaneously capturing skeletal movement as well as high-definition surface physical characteristics in terms of 3-D surface geometry, color and lighting — and literally "imports" this performance into the digital realm, all in



An array of digital cameras captures the actor's image at 24-120 fps. The cameras are positioned to capture the actor from a variety of angles, enabling determination of the surface geometry with extreme precision.

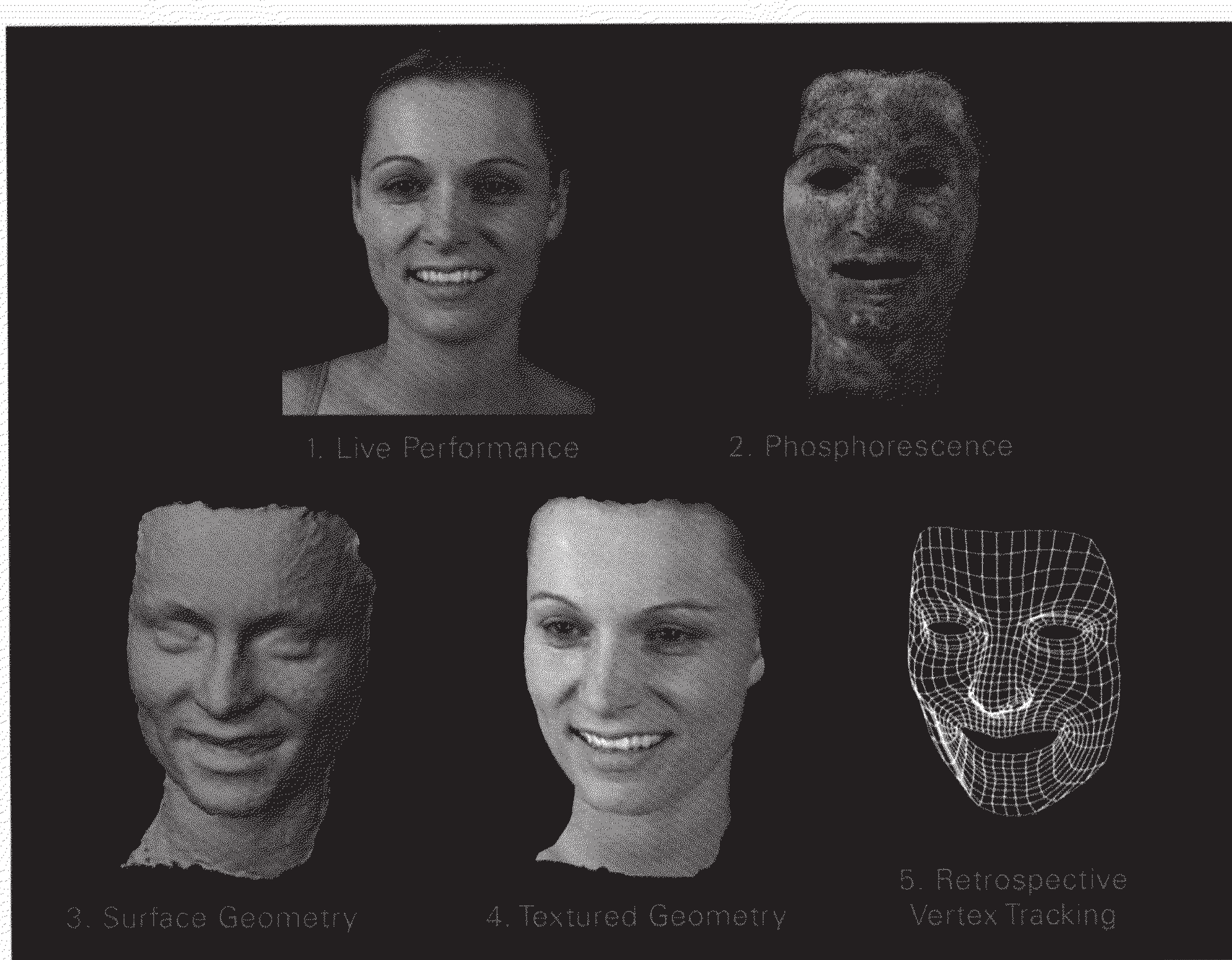
a single real-time pass.

Contour records the image with high-resolution digital cameras, and the subject may be lit for any desired effect — for instance, to match a previously determined background or setting. The cameras can be run from a standard 24 fps up to 120 fps, allowing for the capture of fast action or the creation of slow-motion effects. The geometric information is recorded by an array of grayscale digital cameras — no less than two and as many as hundreds — with each photographing the performers from a slightly different angle. The visual information is recorded by a smaller array of color digital cameras — no less than one and as many as dozens — also photographing the performers from a variety of angles.

Contour is used on a light-tight soundstage lit by custom stroboscopic fluorescent lamps that illuminate the subject with white visible lights and black lights that are synced with both camera systems. Typically strobing at 90 times per second, the flashing effect is invisible to the human eye, which perceives the scene as normally lit. The visual-camera shutters open each 1/180th of a second and capture information while the fluorescent lights flash on (these are the "lit frames"). The geometry-camera shutters open each 1/180th of a second while the fluorescent lights flash off and the soundstage is in pitch darkness (the "dark frames"). And what these geometry cameras see in pitch darkness is the most interesting part of the Contour system.

A special, FDA-approved, hypoallergenic phosphorescent makeup (similar to the glow-in-the-dark makeup sold in Halloween stores) is applied in random patterns onto everything in the scene that is to be captured, including an actor's lips and nostrils and right up to the edge of the eyes. Charged during each "lit" frame, the phosphor emits a random-pattern afterglow during the "dark" frames. These patterns are captured by each geometry camera from a different angle and then fed into a small array of computers. The computers correlate the random patterns seen by the cameras from different angles and triangulate between the cameras, producing a 3-D model of every surface visible in the scene during that frame. Anything that is not covered by the phosphorescent coating is not captured.

Using 1.3-megapixel cameras, Contour can reconstruct the 3-D surface of a human face with better than submillimeter precision, resulting in a high-resolution 3-D mesh in excess of 100,000 polygons. This is more than enough resolution to pick up the shape and motion of wrinkles and nostril flares. In comparison, the faces in the mo-cap feature *Beowulf* were captured with fewer than 200 polygons using a marker-based system, and it was then up to a post team to synthetically create



The five stages of the Contour process: 1) video of a live performance is captured during each "lit" interval; 2) a phosphorescent image is captured during each subsequent "dark" interval; 3) data from the phosphorescent images is used to construct high-resolution 3-D surface geometry of the actor; 4) to provide an onset preview, textured geometry is created by applying the live performance image onto the 3-D surface; 5) finally, a vertex tracking mesh is generated to allow animators and effects artists to retrospectively define the vertices they need to track.

all the details in the faces.

Contour's resolution is so far in excess of what is needed to achieve photo-realistic results that in many applications, lower-resolution meshes (e.g., 1,000-2,000 polygons) are preferable. To this end, Contour offers a feature that allows users to retrospectively select only the points on the face where polygon vertices are required for their specific application, whether it be for high-resolution feature-film use or a video-game character. The system will then produce a lower-resolution mesh that precisely tracks these vertices through the performance from frame to frame. This also allows a single captured performance to be stored as a permanent asset that can be repurposed any number of times, saving both talent and production expenses.

The phosphorescent makeup is mostly invisible under normal lighting conditions, so by combining the phosphor with appropriate base makeup, filmmakers can achieve almost natural skin tones. The performer's eyes and teeth — not covered by the makeup — can be tracked optically.

After a Contour mo-cap shoot, the surface geometry and the visual image are "reconstructed" overnight (each frame currently takes less than 60 seconds to compute), resulting in a high-res, full-motion, naturalistic 3-D representation of the subject, which can then be easily manipulated with such animation programs as Autodesk's Maya, 3ds Max and MotionBuilder, or Avid Softimage's XSI and Face Robot. "We've designed Contour to work with as many tools as possible," notes Perlman. "In fact, there's a $250 piece of software called Poser that is used to pose 3-D characters, and we've been dropping data right out of Contour and into Poser without a hitch."

During the four-year development of Contour, the use of phosphorescent makeup arose as a novel solution to a distinct dilemma. "We knew we had to have something that could take a relatively smooth surface and give us a texture that we could grab hold of and track," says Perlman. "The hard thing is that we're often dealing with actresses who invest a lot of time and energy into making their skin perfectly smooth. If it's perfectly smooth, then there's no way to determine the shape of the face. So we tried several different things.

"We first tried putting black paint on the face and capturing the reflection of light on the surface. But we learned that when you're dealing with reflected light, each camera is going to see the reflection differently, depending on its point of view of the surface. That's why marker-based capture uses retroflective markers — they reflect back to the point of the light source. So as the character moves, you see a nice bright spot on the camera because the lights are around the camera lenses. If you use reflection as opposed to retroflection, you get all screwed up, because as the character moves, any highlights at the point of reflection

move, making it look like the face is moving around even if it isn't. So we couldn't get reliable results. One solution was to use extremely flat lighting, where you put the subject in a lock-down situation and use reflector boards all around to eliminate highlights. You can get good capture results by doing that, but it's just not practical when you have actors walking around and you need dramatic lighting. They may sweat and have shiny spots on their faces and so forth."

Perlman and his team then considered using retroreflective paint, "which is used for things like highway-safety signs. But it's not safe to apply that stuff to human skin, and also, it doesn't deform. Retroreflective paint has tiny glass beads in it that reflect light back to the source. First of all, this would be dangerous to ingest or get in your eye, and second, the paints are really rigid once they dry because of the glass beads. We needed something that would stretch and move with the skin. We also tried ultraviolet paints to see if we could get a good capture. It would work in controlled situations, but it wouldn't work easily and efficiently in a typical production situation.

"By that point, we were running Mova, and we were extremely sensitive to the costs and challenges associated with production. So we started looking at phosphorescent pigments and the notion of flashing the lights on and off. This was based on knowledge I'd gained when I designed large-screen monitors for Apple and researched the human visual system's threshold for flicker perception. You're much more sensitive to flicker perception in your peripheral vision than in your foveal [center-view] vision. Large monitors have more of their area in your peripheral vision, so running large monitors at 60Hz, which is where Apple was running their small monitors, would create an annoying flicker in the corner of your eye. So we started testing monitors, running them at up to 80 or 90Hz, and I was able to determine at what point humans stop seeing flicker.

"Twenty years later, we were able to apply that knowledge to the strobing lights we use for Contour, making them flash on and off at a rate that is imperceptible to the human eye. Lo and behold, we can sync the camera-array system so that the shutters are only open when the lights are dark and only see the phosphor of the makeup. Because phosphor is emissive rather than reflective or retroflective, we can get a clean read without highlights or shadows."

Of the lighting setup used with Contour, Perlman says, "The black lights do a good job of charging the phosphor in the makeup and are largely invisible to the RGB-color cameras used to capture the visual image of the subject. The lighting units are modified Kino Flo motion-picture fixtures fitted with both black-light and white-light tubes. Each holds four tubes, and we tend to mix them up to even out the illumination. Black lights are also placed all around the stage to evenly illuminate and charge the phosphor. The white lights are placed however you want to light the subject from the standpoint of beauty lighting. For example, if you want the subject to be in a half-light shadow, you would have the combined white and black lights on the illuminated side, and only black lights on the "dark" side. So the white light defines the normal, visible lighting, while the blacklights are there strictly to evenly charge up the phosphor. All the lights strobe during the capture process, with the phosphor glowing during the dark phases. This is all done in a darkened studio. Also, there is a key difference between fluorescence, the glowing illumination that occurs when phosphor is exposed to light, and phosphorescence, which is the afterglow. We're relying on phosphorescence."

Perlman notes that because the raised reference markers placed on the performer's face for traditional mo-cap work are slightly offset from the surface of the skin, the resulting data is not always a precise representation of how that surface moves. By having the random phosphorescent patterns applied directly on the skin, Contour captures the geometry of the surface itself. By combining this detailed information with skeletal motion, Contour can also be used in conjunction with a marker-based system, allowing filmmakers to use the best of both technologies.

The only special requirement placed on the performer is that he or she must wear the phosphorescent makeup, which is mixed with a standard base and applied with a sponge like regular makeup. The subject must not touch or otherwise disturb the makeup once it is applied. Although a "smudge" will not alter the surface geometry as perceived by Contour, it will result in a discontinuity of any retrospectively tracked vertex that falls within the smudge. The system would continue to capture the performance with full 3-D resolution, but some post cleanup of the data would be necessary to link the pre-smudge vertex location with the post-smudge vertex location. Some post tweaking might also be needed for tracking vertices in areas of very high surface deformation, such as

around the lips and eyes. (Mova expects that future versions of Contour will automate more of this process.)

Contour allows for multiple actors to be captured simultaneously, allowing for complex group scenes. The performers' hand movements and gestures can also be captured. Contour relies on the ability of its cameras to have an unobstructed view of a given surface; if a hand is holding an object and only the outer surface of the hand is visible to the cameras, then Contour will only be able to reconstruct the outer surface.

Contour was designed to offer users on-set creative control that is close to what they expect with traditional cinematography. Although it typically takes an overnight render to reconstruct a full-resolution capture session, the system provides a low-res preview version of the finished 3-D reconstruction image that can be quickly generated on-set, allowing the cinematographer to properly light the subject for a desired effect and the director to check the performance. Reducing the amount of time required to create full-resolution imagery is simply a matter of applying more computing power to the processing. Full-resolution images in real time on set are possible with Contour today at a cost that could be justified by a large-scale production, but within a few years, given the steady advance of computing power, it will be within the reach of more modest production budgets.

Contour can also be adapted for use on any moving camera platform, even a Steadicam or Helmetcam, giving the user great creative latitude.

The system can even be used to capture the geometry and textures of textiles in motion. Clothing is time-consuming and difficult to faithfully simulate with CG animation techniques, but Contour can capture a garment's exact geometry, motion and texture once it has been treated with a phosphor-based dye. Rather than donning the form-fitting Lycra suits necessary for marker-based mo-cap work, actors can be costumed for their characters, and Contour will capture the fabric's motion and the actor's motion simultaneously.

Perlman also foresees using Contour in conjunction with traditional stop-motion animation. "You just mix the phosphor in with the material being used, such as clay or silicone," he explains. "*The Corpse Bride* was done with puppets made of silicone over metal armatures, while *Wallace & Gromit: The Curse of the Were-Rabbit* was done with modeling clay. You would mix the phosphor in at a low-enough density that you wouldn't notice it under normal stage lighting — you'd just see the normal colors of the silicone or the clay. But when you turn off the lights, it will glow, and because the phosphor is just a powder that is mixed in, there will be a random pattern to it.

"A complex shot like Wallace and Gromit driving through people's gardens in a car is very expensive, because it contains so many different objects that must all be animated," he continues. "You also have a motion-controlled camera that's following the action. All this is very difficult to coordinate and execute. An alternative might be shooting the scene with Contour and breaking it into components on separate sets, lighting and animating them as you would normally, and then combining the images and rendering them in 3-D on a hi-def monitor for viewing. Because it's now a digital version of the physical puppets, you would be able to composite characters and elements together using a tool like Maya for evaluation. In this case, Contour could be used to replace a motion-controlled camera, because once the digital information was captured, you would have full freedom of motion in 3-D; you could zoom in, pull back or fly through the air. And if you have a character that flies, rather than being forced to suspend the character with some sort of harness, you could just put it on a separate set and position it on the ground in the position you want it to be in when it's flying.

"Also, if you wanted to reduce the audience's perception of the strobing effect of stop-motion animation — or not, if that's part of your artistic expression — you could just turn on motion blur in your 3-D package and instantly turn stop-motion into go-motion. So we think Contour could also dramatically lower the cost of stop-motion animation and afford a lot more creative control.

"In the parlance of computer graphics, Contour is the first technology to successfully cross the 'Uncanny Valley,' a perceptual zone where a CG face looks almost photo-real, but not quite photo-real. Such images are disturbing to the human visual system because our brain thinks it's seeing a face with some defect. Humans don't have that reaction to caricatures, such as cartoon faces, because we know they aren't real. But as faces approach photo-real, either you are spot-on, or you have something worse than a caricature. Contour will give you a spot-on photo-real face. Then you can focus your energy on what you should be worrying about: making a great movie or video game."

For more information and to view demos, visit www.movapodcast.com.